

Bezpieczeństwo danych w internetowych sieciach społecznych na przykładzie portalu nasza-klasa.pl

Artur Strzelecki, Tomasz Bacewicz, Marcin Ściański

Wprowadzenie

Analiza sieci społecznych jako nauka, wspiera nas w badaniu i ocenianiu potencjału sieci, oraz portali społecznościowych rozwijających się w Internecie. Jako członkowie społeczeństwa informacyjnego nie możemy przejść obojętnie obok ogromu danych i informacji jakimi dysponują sieci tego typu.

Sieci społeczne w Internecie

Teoria analizy sieci społecznych znajduje swoje początki w teorii grafów. Sieć to struktura złożona z węzłów oraz jednego lub więcej typów relacji je łączących. Finalny obraz sieci społecznej często jest bardzo rozbudowany, wynika to z bezpośredniego przeniesienia realnych złożoności na symboliczną strukturę grafu.

Sieci społeczne zarówno te scentralizowane jak i rozproszone oparte o Internet święcą obecnie triumfy. Ilość użytkowników, ich aktywność oraz poziom zaufania jakim obdarzają portale jest niezmiernie wysoka. Zasób informacji jakie użytkownicy przekazują do tych systemów jest bardzo obszerny. Niektóre z platform stały się wręcz ogromnymi systemami stworzonymi zgodnie z ideą Web 2.0 dla których tak jak Facebook

można tworzyć dedykowane aplikacje webowe, dzięki specjalnie zaprojektowanemu na potrzeby portalu interfejsowi API. Portal taki jest bazą do rozwoju związanego z nim biznesu przeznaczony dla dokładnie wyselekcjonowanej grupy użytkowników/konsumentów. Co więcej, portal ten stał się platformą marketingową, zdolną kreować nowe potrzeby konsumentów, w pełni wykorzystując możliwości szybkiego dostępu do informacji. Użytkownicy portalu otrzymują oferty reklamowe stworzone i wyselekcjonowane na podstawie oceny preferencji użytkownika korzystającego z witryny. Z pomocą administratorom przychodzą rozbudowane systemy statystyk, oraz informacje zapisywane na komputerze w postaci *cookies*, zbierają one godziny oraz dni tygodnia w jakich użytkownik korzystał ze strony, oraz jakie funkcjonalności są jego ulubionymi.

Wspomniane Web 2.0 pozwoliło na silny rozwój platform społecznych takich jak Facebook czy MySpace. Web 2.0 to ogólne pojęcie wyróżniające osiem najważniejszych cech nowoczesnych portali internetowych:

- możliwość nawiązywania kontaktów (ang. *connectedness*) - zapewnia istotę sieci społecznej, pozwala na poszerzanie kontaktów interpersonalnych jak i biznesowych,
- łamanie istniejących zasad (ang. *shattering the existing*) – wyznacza nowoczesny rozwojowy trend dla aplikacji webowych,
- partycypacja (ang. *sharing*) – wymaga od użytkownika zaangażowania w tworzeniu portalu,
- kreatywność (ang. *creativity*),
- niskie koszty (ang. *low cost*) – w ostatnich 25 latach koszt stworzenia portalu społecznościowego spadł z 20 mln USD do 2 mln USD,

- to czego chcę i kiedy chcę (ang. what I want) – czyli możliwość dostosowania interfejsu, do osobistych potrzeb,
- "śmiertelność" (ang. death) - zanikanie w warunkach konkurencji witryn których tematyka jest związana z przemijającym już trendem,
- szybkość (ang. speed) – portal szybko reaguje na zmiany rynku i wymagania konsumentów [Nied07].

Wymagania jakie stoją przed witrynami społecznościowymi to przede wszystkim wysoka konkurencyjność i dynamika w podążaniu za potrzebami rynku. Skutecznie działająca witryna, musi wykorzystywać najnowsze dostępne technologie, a jej interfejs powinien spełniać wszystkie normy związane z użytecznością. Twórcy portalu muszą znaleźć wspólny dla wszystkich temat, będący w stanie połączyć dużą grupę ludzi i dający im jasne korzyści. Kwestią często nie rozważoną przez wielu użytkowników portali społecznych pozostaje bezpieczeństwo danych które sami do niego przekazali, oraz możliwość wykorzystania ich do celów innych niż zamierzone dla danego portalu. Na przykład wysyłanie spam'u na adresy e-mail przekazane portalowi przez użytkowników.

Portal społecznościowy Nasza-Klasa.pl

Historia portalu społecznościowego „nasza-klasa.pl” sięga lipca 2006 roku, kiedy została zarejestrowana domena. Sam portal został uruchomiony 11 listopada 2006 roku. Twórcami serwisu są wrocławscy studenci, którzy adoptowali istniejący pomysł na potrzeby polskich realiów. Protoplastą rodzimego serwisu był amerykański „classmates.com”, który powstał już w 1995 roku i położył fundamenty pod to, co dziś nazywamy „sieciowymi społecznościami” i co jest podstawą idei „Web 2.0”. Za jego

pomocą zarejestrowani użytkownicy mogą wyszukiwać swoich dawnych szkolnych znajomych. Co jednak odróżnia dwa wspomniane portale, to główne źródła finansowania. Podczas gdy „nasza-klasa.pl” jest serwisem zupełnie darmowym, czerpiącym zyski w znakomitej większości dzięki reklamodawcom, „classmates.com” pobiera od użytkowników abonament. Jest to jedna z przyczyn braku zainteresowania serwisem i klęski oferty publicznej portalu. Emisja papierów wartościowych przyniosła straty, wskazując na sceptyczne podejście inwestorów co do przyszłości amerykańskiego serwisu.

Z drugiej strony można zauważyć ogromny sukces „naszej-klasy”. Dane z badania Megapanel PBI/Gemius przeprowadzonego w czerwcu 2007 roku mówią o 497 tysiącach zarejestrowanych użytkowników i zasięgu 3,61%. Wzrost liczby korzystających z serwisu w tempie 8000 rejestracji dziennie doprowadził do tego, iż obecnie już jedna czwarta Polaków posiada swój wirtualny profil. Wśród czynników wpływających na niespotykaną dotąd w polskim Internecie popularność witryny, największe znaczenie ma z pewnością brak opłat pobieranych od użytkowników. Twórcy stopniowo dodają coraz to nowe funkcjonalności. Specjaliści podkreślają również aspekt psychologiczny całego zjawiska. Powrót do przeszłości i odnajdywanie dawnych znajomych wywołują silne emocje, budzą w ludziach sentyment. Dla osób, które założyły już własne rodziny w wielu przypadkach portal stał się jedyną szansą na nawiązanie kontaktu ze szkolnymi przyjaciółmi. Natomiast dla ludzi młodych „nasza-klasa” może być sposobem na wyrażenie własnej osobowości oraz zaprezentowanie siebie, na przykład poprzez publikację fotografii [InSt08].

Opisując w skrócie działanie portalu, można powiedzieć, że jego główną i najczęściej wykorzystywaną funkcjonalnością jest możliwość zapisywania się do wirtualnych szkół i klas oraz wyszukiwanie osób. Użytkownik może identyfikować wyszukiwane osoby, dzięki wspomnianym wcześniej fotografiom, ale w szczególności, dzięki metryczce, zawierającej podstawowe dane osobowe. Do tych danych należą:

- imię i nazwisko oraz nazwisko rodowe,
- pseudonim szkolny,
- telefon,
- miejscowość,
- wiek,
- płeć,
- numery: gadu-gadu i skype.

Kontakt pomiędzy użytkownikami zapewnia skrzynka pocztowa, umożliwiająca nieodpłatne wysyłanie oraz odbieranie wiadomości tekstowych. Aby dodatkowo usprawnić bezpośredni kontakt, na początku kwietnia wprowadzono „nkfon”, pozwalający łączyć się poprzez telefon stacjonarny z wybranymi krajami po preferencyjnych taryfach. Usługa wykorzystuje coraz popularniejszą technologię VoIP i jest dostępna dla abonentów wszystkich większych operatorów telefonicznych w Polsce.

Wszystkie funkcjonalności, mające usprawnić korzystanie z portalu nie odwróciły uwagi opinii publicznej od kwestii bezpieczeństwa umieszczanych w serwisie danych. Wprawdzie na podstawie ustawy o ochronie danych osobowych informacją pozwalającą jednoznacznie zidentyfikować osobę jest numer PESEL (który podlega ochronie i nie występuje w danych osobowych użytkowników portalu), to kombinacja danych, takich

jak na przykład imię, nazwisko, miejsce zamieszkania i telefon daje spore możliwości identyfikacji pojedynczej osoby. Dodatkowo, o czym można było usłyszeć w mediach, zamieszczane są opisy profilu zawierające informacje o wykonywanej profesji czy statusie społecznym, co może stanowić zachętę dla włamywaczy, nie tylko tych komputerowych.

Faktycznie, co jakiś czas na stronie głównej „naszej-klasy” pojawiają się prośby o ignorowanie określonych maili lub programów rzekomo przyspieszających działanie portalu. W rzeczywistości programy te mogą ułatwić przejęcie konta lub komputera i umożliwić uzyskanie dostępu do prywatnych danych użytkownika. Słabość portalu również została obnażona na początku bieżącego roku, kiedy to w Internecie pojawiła się informacja o uzyskaniu dostępu do danych osobowych setek tysięcy użytkowników. Za narzędzie posłużyła powszechnie dostępna biblioteka PHP i podstawowa znajomość jej użycia (rys 1.).

	A	B	C	D	E	F	G	H	I	J	K	
1	id user	id photo	name	surname	rodowe	pseude	tel	city	age	sex	aa	si
12451	7000001	7000001	Ród			Zaprasza Piotr Sosinka		Caly świat	26	?		
12452	7000002	7000002	Anna					Kcynia	42	Kobieta		
12453	7000003		Agnieszka			?			26	Kobieta		
12454	7000004	7000004	Rafal					Biale Blota / Bydgoszcz	18	Męczyzn		
12455	7000005		Grazyna					Gdańsk	53			
12456	7000006	7000006	Agata			Manczkin		Wadowice	37			
12457	7000007	7000007										
12458	7000008	7000008	Marek			kempes helmut		szczecin	45			
12459	7000009	7000009	Ewelinka					Lipce Reymontowskie	13			
12460												
12461												
12462												
12463												
12464												
12465												
12466												
12467												
12468												
12469												
12470												
12471												
12472												
12473												
12474												
12475												
12476												
12477												
12478												
12479												

Rys 1. Plik xls z danymi użytkowników portalu „nasza-klasa” [HACK08]

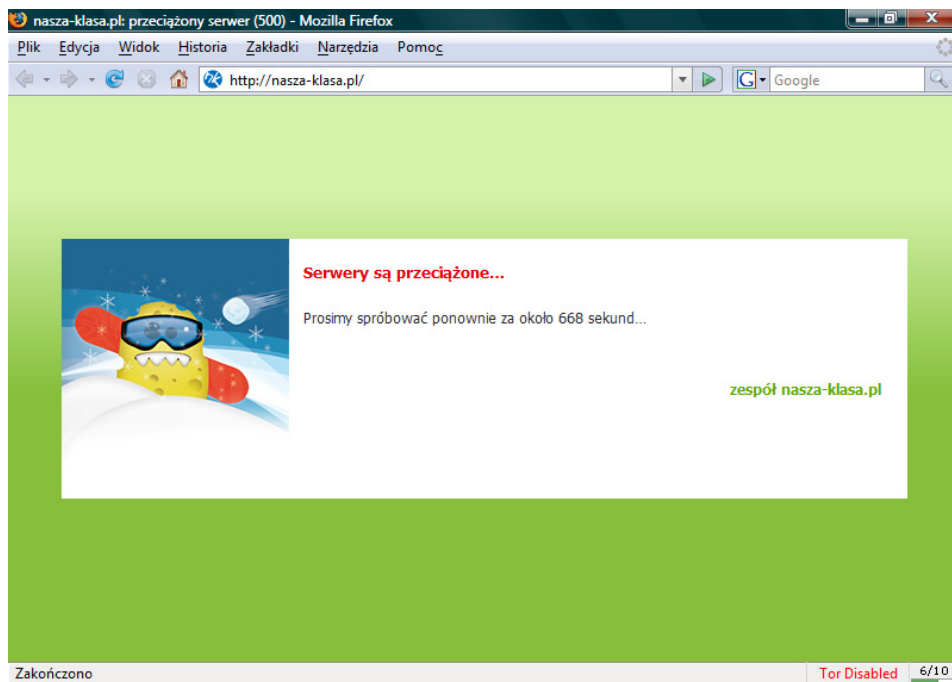
Znalezienie inwestora zagranicznego dało nadzieje na uporanie się z problemem bezpieczeństwa. Równie ważna dla przeciętnego użytkownika była także szybkość działania portalu, która pozostawiała wiele do życzenia. Bardzo często zamiast ekranu logowania można było zobaczyć informację o przeciążonych serwerach. Z upływem czasu portal działał coraz szybciej, a administratorzy publikowali informacje o kolejnych zabezpieczeniach mających chronić dane. Jak się jednak okazuje cały czas istnieje możliwość pobrania danych osobowych dużej liczby użytkowników portalu.

Pozyskiwanie danych z Nasza-Klasa.pl

Motywacją do przeprowadzenia badań była chęć pozyskania danych, które zgodnie z zapowiedziami właścicieli tego portalu są bezpieczne. Autorzy pozyskali próbkę danych do badań poprzez zbudowanie robota sieciowego. Przeszukiwał on strony internetowe zawierające dane personalne zarejestrowanych użytkowników portalu. Autorzy posiadają już pewne doświadczenie przy pozyskiwaniu danych z heterogenicznej sieci web. We wcześniejszej pracy przeprowadzili badanie pobierając wszystkie dane o zrealizowanych aukcjach internetowych w serwisie aukcyjnym Allegro.pl [StBaSc08]. Robot sieciowy został od nowa zbudowany z wykorzystaniem zupełnie innych modułów w języku Perl niż wykorzystano w poprzedniej pracy. Wynikało to potrzeby logowania się poprzez nazwę użytkownika i hasło, aby móc zobaczyć dane, które miałyby zostać pobrane. Cały kod robota sieciowego został napisany w Perlu, natomiast dane „w locie” były zapisywane do bazy danych MySQL.

Pierwsza część w kodzie źródłowym była odpowiedzialna za nawiązanie połączenia, zapisanie sesji oraz utrzymywanie stale zalogowanego wirtualnego użytkownika, który pobierał kolejne strony portalu do analizy. Druga, ważniejsza część kodu źródłowego była przeznaczona do wyłuskania za pomocą wyrażeń regularnych istotnych informacji o zarejestrowanych użytkownikach. Robot pobierał i zapisywał do bazy danych następujące informacje: imię, nazwisko, pseudonim szkolny, nazwisko rodowe (jeśli występowało), telefon, miasto, wiek, płeć, numer w komunikatorze Gadu-gadu oraz nazwę użytkownika w komunikatorze Skype.

W trakcie kilku prób uruchomienia robota okazało się, że portal nasza-klasa.pl nie pozwala automatycznie przeglądać większej ilości stron jednemu użytkownikowi. Podjęto próby zbliżenia zachowania robota do bardziej naturalnego dla zwykłych użytkowników, jak wyposażenie robota w odpowiednie wartości User-Agent oraz Referer. Niestety wydłużyło to tylko nieznacznie moment, w którym nasz wirtualny użytkownik o nazwie „pająk informatyczny” był blokowany przez portal. W tym krótkim czasie około 15 minut robot był w stanie pobrać na łączu o prędkości 512kbps informację o jednym tysiącu użytkowników. Nie ulega wątpliwości, że to działanie zabezpieczające zostało podjęte z uwagi na liczne już próby pobrania danych z portalu. Z pewnością świadczy o tym również komunikat (rys. 2) jaki zostaje wyświetlony po zablokowaniu konta z powodu nadmiaru pobranych danych.



Rys 2. Komunikat zawiera następujący opis: Pan Gąbka dostaje kulkę
źródło: [NK08]

Podjęto próbę oszukania serwera dostępowego poprzez cykliczną, częstą zmianę numeru IP komputera, na którym działał robot. Doskonale do tego celu nadaje się The Onion Routig (trasowanie cebulowe) [TOR08]. W sieci Tor tworząc ścieżkę prywatną, oprogramowanie użytkownika i klienci stopniowo budują obwód szyfrowanych połączeń pomiędzy przekaźnikami sieci. Obwód jest rozszerzany za każdym razem o jeden węzeł, tak że każdy z przekaźników po drodze zna tylko przekaźnik, z którego otrzymał dane i przekaźnik, któremu dane wysyła. W ten sposób robot sieciowy bez wykrycia działał wiele godzin i stale pobierał dane. Ponieważ jest to rozwiązanie niekomercyjne, nie było ono do końca stabilne. Po kilku godzinach pojawiał się w sieci Tor na tyle długi timeout, iż pobieranie danych przerywało się z powodu wygaśnięcia sesji wir-

tualnego klienta. Do bazy została pobrana próbka danych wielkości 5022 rekordów. W tej grupie ok. 10% użytkowników podało swój numer komunikatora Gadu-gadu, jednak już tylko mniej niż 2% użytkowników podało swoją nazwę w komunikatorze Skype. Numer telefonu był dostępny u około 2% użytkowników.

Powyższe wyliczenia pokazują, że użytkownicy pragnąc zachować pewien określony stopień prywatności, nie podają informacji umożliwiających bardzo łatwy, bezpośredni kontakt. Pomimo euforii panującej wokół portalu, użytkownicy zdają się rozumieć związane z nim zagrożenia.

Korzystając z portali społecznościowych takich jak Nasza Klasa musimy pamiętać że dane umieszczone na takim portalu są łatwo dostępne. Przy użyciu paru wyspecjalizowanych metod mogą zostać pobrane i wykorzystane na przykład do rozsyłania spamu. Bezpieczeństwo naszych danych jest więc zależne przede wszystkim od nas i naszej ostrożności. Jeśli to tylko możliwe powinniśmy podawać jedynie podstawowe dane, konieczne do odnalezienia nas na danej witrynie.

Literatura

- [HACK08] <http://hacking.pl>, dostęp 27.04.2008 r.
- [InSt08] <http://www.internetstandard.pl>, dostęp 27.04.2008 r.
- [NK08] <http://www.nasza-klasa.pl>, dostęp 27.04.2008 r.
- [Nied07] Niedbała M.: Budowanie społeczności internetowej Dizajnerzy.pl z wykorzystaniem technologii Web 2.0, *Metody informatyki stosowanej*, Nr 1/2007 (Tom 11), s. 85-90

- [StBaSc] Strzelecki A., Bacewicz T., Ściański M.: Ocena zachowania użytkowników platformy handlu C2C, e-mentor 2008 numer 2(24), s.76-82
- [TOR08] <http://tor.eff.org>, dostęp 27.04.2008 r.

Informacje o autorach

Mgr Artur Strzelecki
Tomasz Bacewicz
Marcin Ściański
Katedra Informatyki
Akademia Ekonomiczna
ul. Bogucicka 3, 40-226 Katowice – Polska
e-mail: strzelecki@ae.katowice.pl
tomek.bacewicz@gmail.com
marcin.scianski@gmail.com